Final Technical report for AFOSR F49620-95-C-0018

Title: A Fidelity Metric for Image Sequences

Period covered: Feb 15, 1995 - May 14, 1999

Objectives of research effort:

The development of a computational model of human vision, which might also be used as an image fidelity metric, involves basic psychophysical research to characterize the mechanisms of early vision. Of particular concern, and the basis of many of our experiments, is the understanding of visual masking particularly as it applies to acuity and other tasks relevant to natural scenes. The approach we employed to examine this problem is the test-pedestal paradigm, a common design of psychophysical studies of visual function which has analogous application to complex images containing artifacts resulting from image compression or other processing technologies. In terms of the test-pedestal paradigm, the image artifacts are considered the test whereas the original image is the pedestal masker. Can we predict performance on masking tasks using simple targets such as line bisection, Vernier acuity, and two-line resolution? Our results demonstrate we can, moreover detection thresholds on complex targets are also readily predicted. However, when applying standard filter based models to complex scenes we find that performance is not readily predicted. Many additional factors not appreciated before this research began needs to be addressed before the successful creation of a comprehensive vision model. Our research has highlighted these problems, which we discuss in detail in our publications.

Overview of the final report:

The past four+ years of research supported by this grant been very productive with significant progress made on several fronts. During this time 31 publications and presentations have completed with AFOSR support and more are expected as the manuscript writing continues. Copies of four published or in press manuscripts, which have not been included in prior technical reports, are included with this report. These papers include additional details of our research effort, beyond that contained in prior quarterly and annual reports. This final report summarizes the research program and describes the projects we have completed. Since much of this material has been presented in prior submitted technical reports we will focus more on our research effort since the last annual report.

The research performed covers a variety of topics but all have been designed to contribute to the underlying goal of extending our ability to model human vision. The better the model the better the image fidelity metric, a metric that evaluates image fidelity is simply a model of human vision. Going into every research project is beyond the scope of this report. We have artificially grouped the investigations into three categories and provide an overview of the research in each category. The three categories are: 1) Technical developments that further psychophysical research in general and address practical issues in designing a vision model. 2) Basic studies of visual processing using simple stimuli with an emphasis on masking. 3) Masking and real world scenes and its relevance to image compression/quality issues. This section includes the creation of the Modelfest group that, in the spirit of collaboration, brings together top researchers interested in modeling human vision and promises to accelerate progress in vision modeling.

In the following sections the relevant publications/presentations supported by this grant are indicated by number (see publication list) rather than using a formal citation format. We view

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0158

Public reporting burden for this collection of information is estimated to average 1 hour per response, inclu ... sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. S... of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Serv... Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperw...

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 3/30/00	3. REPORT TYPE AND DATES COVERED Final Tech. 2/15/95- 5/14/99

4. TITLE AND SUBTITLE

A Fidelity Metric for Image Sequences

5. FUNDING NUMBERS

F49620-95-C-0018

6. AUTHOR(S)

Thom Carney

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Neurometrics Institute
2400 Bancroft Way
Berkeley, CA 94704

8. PERFORMING ORGANIZATION REPORT NUMBER

FT

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
110 Duncan Ave., Suite B115
Bolling AFB, DC, 20332-8080 NL

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The development of a computational model of human vision, which might also be used as an image fidelity metric, requires basic psychophysical research to characterize the mechanisms of early vision, with special emphasis on visual masking. We have performed a variety of psychophysical experiments using the test-pedestal paradigm to study visual acuity and motion discrimination performance. Performance using simple targets can often be predicted from the observers own contrast sensitivity. The mystery of how humans achieved hyperacuity performance on many spatial vision tasks has essentially been solved. However, performance on simple psychophysical tasks appears to have limited application to detection of a target in complex backgrounds, typical of video fidelity assessment tasks. Several studies indicate the visual system can use adaptive template mechanisms in complex tasks, which are not readily modeled using the fixed filter properties of current early vision models. While masking by contrast gain control mechanisms may be important in simple stimulus known exactly tasks, its roll is less significant in video quality where artifacts are to be discriminated from an unknown background pedestal. Future work should focus on the adaptive nature of visual mechanisms and their task dependent properties.

14. SUBJECT TERMS

visual masking, resolution, spatio-temporal processi

15. NUMBER OF PAGES

21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	UL

NSN 7540-01-280-5500

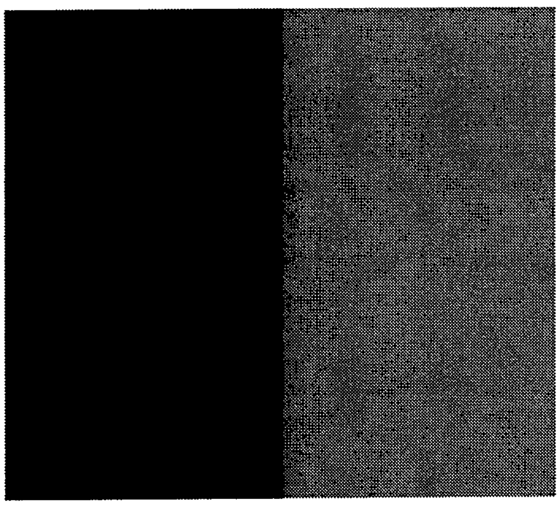
Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

this report as an opportunity to not just summarize our research effort but to also suggest future research directions based on our experiences of the past 4 years.

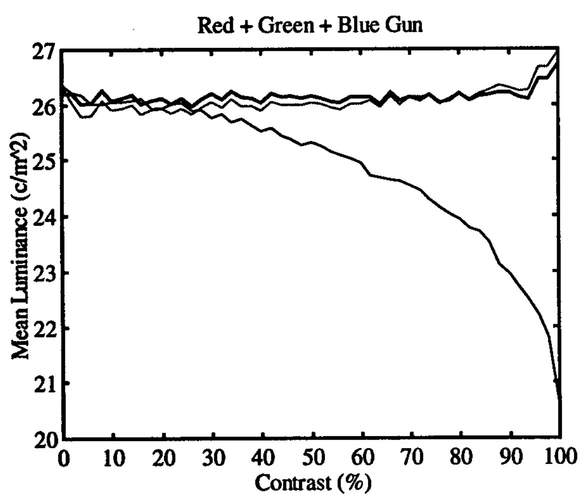
1. Technical developments of general contribution to vision science and vision modeling:

Pre visual system nonlinearity (references #1, 7, & 10):

The application of a vision model to images presented on a video display requires the first stage of a vision model apply luminance compensation to correct for the display gamma function nonlinearity. This stage can be bypassed if, as is common practice in psychophysical experiments, a 1D look up table has been used to correct for the display gamma nonlinearity. However, in the case of natural scenes or most any stimulus with vertical and horizontal features a second display nonlinearity is present that is also most never corrected (many psychophysical studies only use horizontally oriented stimuli for this reason) which we call the adjacent pixel nonlinearity. Video monitors exhibit large adjacent pixel interactions along the raster direction, which changes the mean luminance and contrast of many patterns. The figure on the left simulates the magnitude of the problem. The left and right sides of the figure are high frequency (light-dark alternations), high contrast, vertical and horizontal gratings, respectively. As viewed on a video monitor at a distance the grating structure is not visible but the mean luminance of the left side of the picture is noticeably lower than that of the right side. The adjacent pixel non-linearity can reduce the local mean luminance patterns by up to 30%, even after normal 1-D gamma correction.

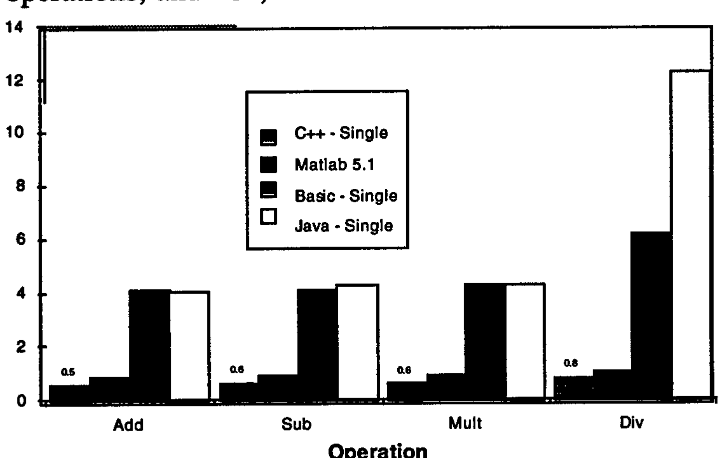


We have devised a novel technology which corrects for the adjacent pixel interactions. The adjacent pixel non-linearity can be modeled using an exponential low-pass temporal filter followed by the monitor's gamma non-linearity stage. The time constant of the low pass filter corresponds to the temporal bandwidth of the video amplifier. We have used this 5 parameter model along with a series of test measurements to develop a two dimensional lookup table (2D LUT) which can be used correct for both sources of luminance error. Our first 2D LUT was limited to a single video color gun, but we have now extended it for use with all three video guns simultaneously as shown in the figure to the left. The stimulus was a vertical grating with alternating light and dark bars. In In the plot, a horizontal line would indicate perfect luminance compensation. The solid line that drops in mean luminance with increasing contrast is for the gamma only correction condition. The solid (dark) horizontal line is the model prediction with gamma and adjacent pixel non-linearity



compensation. The dashed (light) data line is the actual display luminance after applying the 2D-LUT. The jagged structure is due to round off errors associated with the 8 bits/pixel hardware limit. The predicted and observed values are within about 1% of each other. The increase in mean luminance at high contrasts is an artifact that can be easily avoided (see Carney and Klein, 1997). With the use of our new 2D-LUT procedure we can be sure that arbitrary 2D luminance profiles will accurately reflect the correct mean pixel luminance values on the display monitor. Alternatively, a vision model can incorporate 2D-LUT compensation at the first stage of digital image processing, when the video image presentation does not include luminance non-linearity compensation (normal video presentation conditions). This is an important technological advance that will be usefull for the design of future psychophysical experiments and inclusion in applied human vision models.

Practicalities of vision modeling – scarcity of computational resources (#15):
Front-end filter based HVS models are computationally intensive. As it turns out the desk top computer available to me (dual Pentium Pro system) was woefully inadequate. Colleagues working on similar problems have systems with over 1000 times the computational power and they still wish they had more. While computers are becoming ever faster and cheaper, I fear this problem is a barrier keeping interested researchers from working on the development of general purpose vision models. While the suitability of different programming tools and languages is important as I discussed at the annual 1997 Optical Society of America and SPIE meetings, they alone will not solve the problem of inadequate computer resources. As I described at the talk, programming language differences can roughly have about a factor of 10 impact on final run times. Another important factor in choosing a language is ease of use, where languages such as Matlab have enormous advantages over lower level tools such as "C". Our own system is based on a combination of Matlab, an efficient high level interpreted language optimized for matrix operations, and "C", for time critical modules. I have benchmarked the computational efficiency of a few common languages, including compiled JAVA, C++, Matlab and compiled Basic on optimized pointwise matrix operations which are critical for filter based modeling. In the figure, the computational time is shown for each language relative to C++ times for double precision operations. I have discussed the findings in previous reports so I will not be labor them again here. Rather, I bring this topic up to suggest a more cooperative approach to computing resources. Yes, programming languages make a difference but hardware is a bigger factor (besides personal talent) separating the progress of one researcher from another. In my experience, most academic laboratories have desktop computer systems connected to the Internet which spend most of their time idle. Distributed computing has become an important concept in recent years and I think we should consider how to apply it to vision modeling to utilize the countless CPU cycles wasted in labs across the country. Front-end filter models are inherently amenable to distributed parallel processing computing, after all they mimic the structure of the visual cortex, the ultimate parallel processing engine. A federal agency



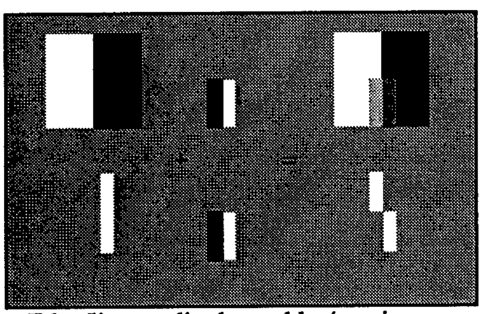
should invest in the development of a mechanism to distribute human vision model computational packets across the internet to registered computers that sit idle much of the time, such as at night. I have explored this approach a little and think it has great promise. The Modelfest group (described below) or some other organization of vision scientists could oversee such a collaborate effort where all who participate stand to benefit without additional cost once a distributed computing vision network is operational.

2) Basic psychophysical studies of visual processing using simple stimuli:

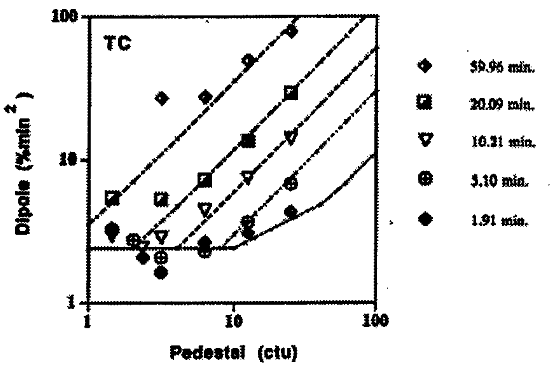
Our general approach throughout this period of research has been to conceptualize diverse perceptual tasks in terms of the test-pedestal paradigm. This often simplifies comparing thresholds across tasks which have the same test but different pedestals and enables us to determine if special mechanisms are required to explain performance beyond those associated with simple contrast discrimination. Using this framework, each experiment generally involves determining the test strength necessary for its detection in the presence of different strength pedestal, or masking, stimuli. On the applied side, this framework has direct utility in developing a fidelity metric where the visibility of image distortions, resulting from the compression and decompression process, is to be identified. In that case the distortion due to compression is the test stimulus and the original image is the pedestal. The question is to what degree does the original image mask the visibility of the test; in this case the compression artifacts. For lack of any other natural distinctions I've categorized the studies into those predominantly using static stimuli and those using dynamic stimuli (the discussions draw heavily from previous technical reports).

Test-pedestal approach with static simple targets: (#9, 6, 16, & 19)

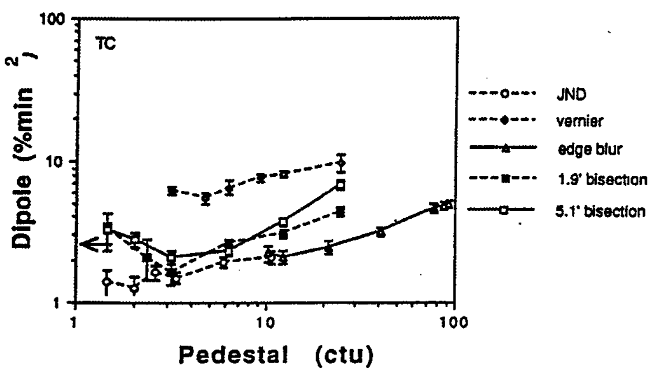
Resolution (blur), Vernier acuity (jaggies) and contrast discrimination (JND) are important aspects image quality. Fortunately, thresholds on these apparently dissimilar tasks can be described in terms of detecting a dipole test stimulus in the presence of a pedestal mask. As the figure to the left shows, adding a dipole to an edge pedestal blurs the edge and adding it to a line pedestal creates a Vernier offset. When adding the dipole to itself it becomes a contrast discrimination task. Analogous combinations can also created using a quadrupole test target with dipole and line pedestals. Since the test was the same in each task thresholds could be directly compared. When thresholds are in dipole test strength units of $\%\text{min}^2$ we find that Vernier acuity is actually worse than resolution and simple contrast discrimination. Hyperacuity tasks such as Vernier acuity no longer appear nearly as impressive now that we understand that performance is no better than what we might expect from contrast sensitivity and may actually be a little worse. We have also used the test-pedestal approach to examine the spatial hyperacuity task, three-line bisection. For three-line bisection, the pedestal is the center line and the test is a dipole which when added to the center line shifts the line to the left or right depending on the polarity of the added test dipole. We have devised a model for predicting an observers bisection threshold as a function of line contrast (pedestal strength) and the separation between the lines of the bisection target. For the data in the figure below the stimulus line separations range from about 2 - 60 minutes. The pedestal strength



Edge/line dipole blur/vernier



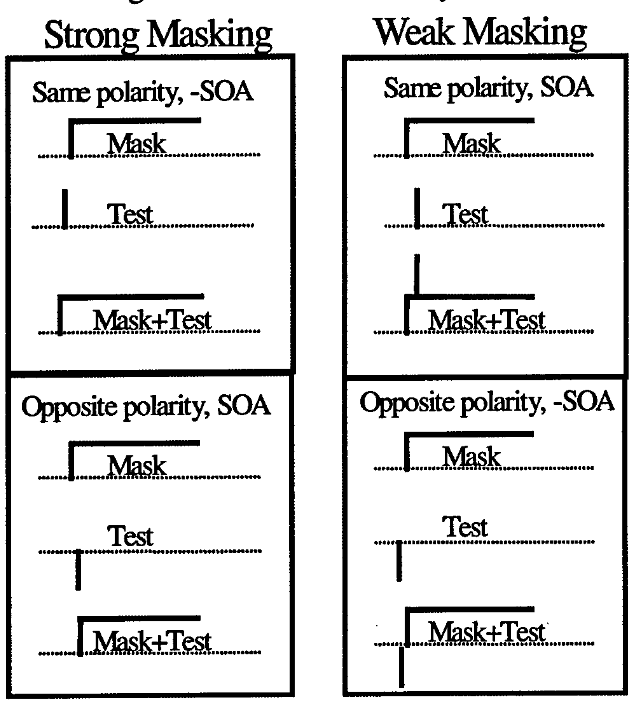
ranged from about 1 to 30 times the line detection threshold (ctu). Thresholds are expressed in terms of the test dipole strength with units of %min^2 (these values can be converted to min of spatial shift if desired). Our threshold predictions for the same spatial separation are shown as dashed lines. The predictions fit the data within a factor of two for all but the lowest pedestal strengths at large separation. The predictions are based on each observer's own dipole *detection* threshold. Threshold is given by the greater of three factors; 1) observers dipole detection threshold, 2) dipole detection threshold * (pedestal strength/10)$^{0.5}$ or 3) line separation * pedestal strength / 60. This formulation captures the idea that three floors limit performance; contrast sensitivity, contrast masking by the pedestal with a slope of 0.5 and finally at large separations a fundamental spatial uncertainty of the visual system, which has become known as the local sign hypothesis. Thresholds can be accurately predicted without need for making the many assumptions of standard filter models.



In the adjacent figure are plotted thresholds for four tasks, three line bisection, contrast discrimination (JND), Vernier acuity and resolution (edge blur), as a function of pedestal strength. The dipole detection threshold is indicated by an arrow along the y axis. Bisection thresholds (open and filled squares) are lower than Vernier acuity thresholds (diamonds) yet above edge blur resolution (triangles) and JND (circles) thresholds. At low pedestal strengths performance on all tasks is within a factor of two of the dipole detection threshold. Masking increases with pedestal strength but at somewhat different rates depending on the task. The bisection task base separations of 1.9 and 5.1 minutes, shown in the figure, bracket the optimal bisection range. In light of the human retinal sampling density, human performance on a diverse set of acuity tasks has long amazed researchers. We now see that performance is actually close to what is predicted from simple contrast discrimination data. Hyperacuity thresholds are actually slightly worse than predicted from JND data. Our research has focused on hyperacuity tasks because they involve high spatial frequency mechanisms. High frequencies have great impact on the efficiency of DCT based image compression algorithms. Where masking is present, especially at high frequencies the compression can be increased without degrading the final image quality. We see from these data that masking under simple stimulus conditions is actually small for the range of strengths tested. Most masking in natural scenes probably involves stimulus uncertainty issues rather than contrast gain control mechanisms, which have become so dominant in the vision science literature (see reference #11).

Test-pedestal approach with dynamic simple targets: (#2, 3, 4, 5, 12, & 13)

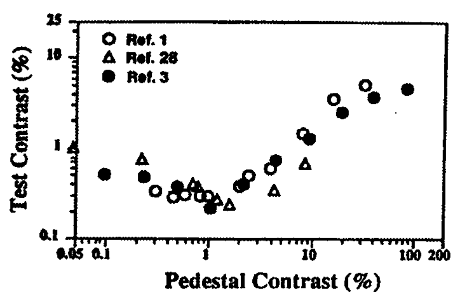
Our studies of visual masking near spatio-temporal edges using Westheimer and Crawford techniques have revealed surprising masking asymmetries that depend on the luminance polarity of the test relative to that of the mask (pedestal). For tests and masks of similar size, when both have the same luminance polarity (light or dark) strongest masking occurs at negative stimulus onset asynchronies (forward masking) whereas for opposite polarity test and mask strongest masking occurs at positive stimulus onset asynchronies (backward masking). We now understand this effect in terms of a stimulus ambiguity that has implications for modeling in general (Strong Masking panel on the left). Implicit in most models of spatial or spatio-temporal vision is the assumption that the stimulus location in space and/or time is know exactly when comparing the system's response to the pedestal alone and the pedestal plus test condition. This assumption is faulty and accounts for some of our unexpected findings. The temporal ambiguity of test-pedestal onset is not treated in any of the current models. The problem also occurs in pure spatial domain modeling, as Al Ahumada pointed out after my SPIE presentation about these results. An image fidelity metric which compares two video streams, the original and the codec version of the original, frame by frame in perfect synchrony, might indicate areas of visible compression artifact which would go unnoticed by a human observer because of this stimulus ambiguity effect. This points to a limitation of test-pedestal approach as applied to real world video streams.



In light of an earlier sections results using static multipole targets, we decided to study blur and resolution using multipoles under motion. Adding a test quadrupole to a pedestal line creates a two-line resolution target. We find that sensitivity to the quadrupole test alone quickly deteriorates with motion when presented alone or in the presence of the pedestal line. Two line resolution performance under motion is predicted by the dipole contrast sensitivity at the same velocities. Edge blur stimuli are generated by adding a test dipole to an edge pedestal. Dipole test detection thresholds also increase rapidly with velocity. However, when added to a strong edge pedestal, blur discrimination performance as a function of velocity is degraded but at a different rate from that of dipole detection. We conclude that the test-pedestal approach offers a simple procedure for evaluating performance on moving resolution tasks in terms of contrast detection sensitivity to the difference signal, the test. Motion deblurring mechanisms appear to offer no special advantage in resolution tasks beyond that of simple contrast detection sensitivity for moving targets. For lines about 10 times detection threshold, two-line resolution acuity is directly predicted by the observer's quadrupole detection threshold over the velocity range tested. Edge

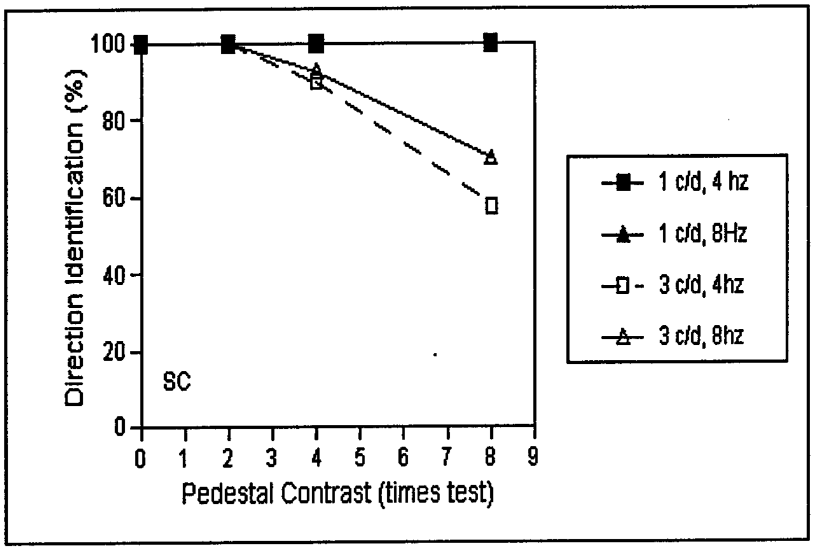
blur sensitivity was worse than the observer's detection threshold for moving dipole test targets. This was expected since the high pedestal strengths used (about 50 times threshold) put us in the Weber masking regime of the test threshold as a function of pedestal strength (TVI) curve for edge blur.

We have also used the test-pedestal approach with sinewave gratings to study grating flicker and oscillatory motion where the test is the same counterphase grating in both cases but added in different temporal phases. Performance was compared for a wide range of pedestal contrast, spatial and temporal frequencies. The main finding was that flicker and oscillatory motion thresholds for supra-threshold sinusoidal gratings are similar, suggesting motion and flicker have a common underlying detection mechanism. The ability to discriminate motion from flicker was elevated relative to their detection thresholds, particularly at high temporal frequencies. We offered two models to account for this behavior. The discrimination of motion from flicker may require a temporal comparison of the outputs of directionally selective filters tuned to opposite directions or to the population statistics of a bank of separable mechanisms. One implication of this study is that the common belief that the motion system saturates at low contrasts, about 2-5% (Nakayama & Silverman, 1985) maybe incorrect. Using this test-pedestal paradigm we observe that when test detection threshold is plotted as a function of pedestal strength the shape is similar to that found for contrast discrimination data. Some facilitation was observed at low pedestal contrasts. At about 10-20 times the test detection threshold a normal Weber-like region was evident. The adjacent figure from our paper (fig 19, reference #13) shows the low contrast motion saturation data from Nakayama & Silverman after transformation into the test-pedestal formalism (open circles). The figure also contains static grating contrast discrimination data from papers by Legge and Foley (filled circles) and Stromeyer and Klein (triangles). When plotted in this way the motion discrimination data look very similar to the static contrast discrimination data. These results support the notion that motion channels may not show low contrast saturation but instead may have a contrast gain control mechanism, or noise, that increases with pedestal contrast. In terms of an image fidelity metric, we now know that the motion mechanisms will not need special saturation masking behavior, they behave just like static contrast detection mechanisms.



There has been a long running debate about whether the early motion system is strictly monocular or includes a binocular component. We have previously shown that when the spatio-temporal quadruture components of a moving grating, which are not themselves moving, are presented dichoptically a moving grating is perceived. Lu and Sperling (Vision Research, 1995) have devised a stimulus decomposition which removes feature-tracking cues that could provide the basis for dichoptic motion perception rather than motion energy detection. The simple addition of a static pedestal grating to each eye's image removes the feature cue and according to Lu and Sperling, abolishes the cyclopean perception of motion. We have subsequently shown that the early motion system is indeed binocular by using a test-pedestal motion stimuli, void of feature tracking cues, which when presented dichoptically elicits the perception of motion. In the figure below, subject SC is able to correctly identify motion direction under dichoptic presentation conditions in the presence of a static pedestal grating which removes the feature

tracking cues. When the spatial frequency was 1 cycle/deg performance was perfect for all pedestal strengths and test temporal frequencies (filled triangle symbols are covered by the filled square symbols).



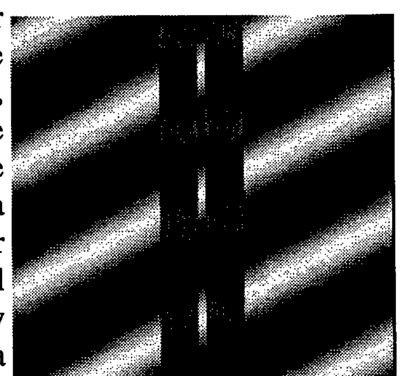
The discrepancy between Lu and Sperling's results and ours involves our use of a longer trial duration so the observer can avoid the masking effect of the pedestal. In addition we have shown that this binocular motion system lacks the low pass temporal frequency behavior which is characteristic of a feature tracking system. Once again the test-pedestal paradigm offers a powerful way of revealing the underlying structure of visual mechanisms.

3. Masking and real world scenes, image compression/quality issues.

This section focuses on more complicated stimuli and the limitations of the traditional approach to the study of masking using simple targets such as those of the previous section. The section has three principal subsections: A) We begin with a discussion of masking using a complex Vernier acuity stimulus and describe models, that predict the results, which diverge from standard filter models. This is followed by studies that applied standard filter models to real world video sequences to evaluate image quality. A new noise masking paradigm is described which enables us to explore mechanisms with adaptable filter properties. B) Next the discussion moves to limitations of the test-pedestal approach and the problems resulting from the fields' emphasis of contrast gain control masking. C) Finally, we end with a review of the activities of the Modefest group, which we have contributed significantly to over the past couple years.

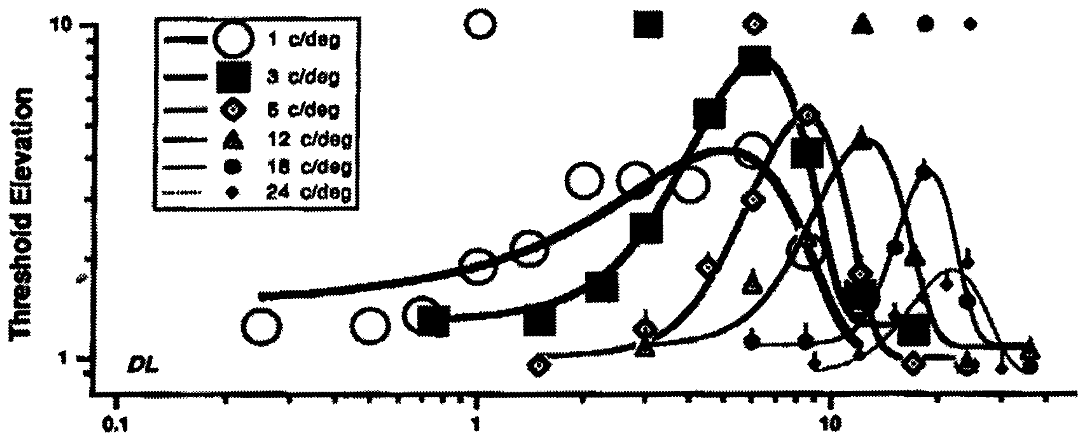
A. Studies of visual masking using complex visual stimuli (# 16, 17, 18, 21, 22, 29, & 31)

We have used the stimulus configuration shown on the right to determine the characteristics of first order mechanisms and their resistance to masking in Vernier acuity configurations.. The Vernier targets are the two narrow central ribbons of grating (static or moving). These ribbon stimuli have two important advantages for studying vernier acuity: 1) they are localized in spatial frequency, and 2) they are localized in there horizontal extent. We measured the orientation, spatial frequency and width tuning of Verneir acuity over a wide range of ribbon spatial frequencies. The results show there are multiple spatial frequency tuned mechanisms which can signal a Vernier offset. For example in the figure below, the vernier thresholds are shown as a function of background mask spatial frequency. Six ribbon spatial frequencies were tested. For low ribbon frequencies (1-5 c/d) the most effective masker was at a higher spatial frequency. These results might offer support to



those who suggest the visual system lacks foveal low spatial frequency mechanisms. Another

striking feature of the data (not shown) is the dependence of frequency tuning on the Vernier ribbon width. The results pose serious problems for current models of early visual processing, they are incompatible with an oriented filter, line element model, in which differential responses of a number of independent filters are pooled across spatial frequency, orientation and space. We performed a wide variety of calculations to determine if filter models could account for the pattern of results obtained. Our modeling shows that threshold elevations predicted by filter

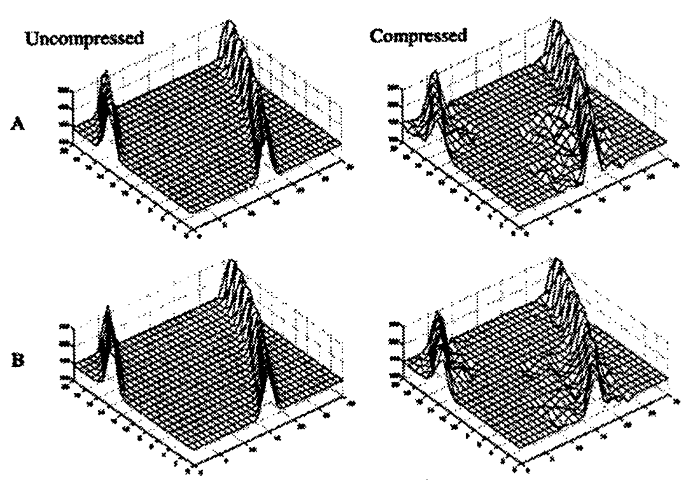


models (with a wide range of filter bandwidths, sensitivities and noise levels) are off by, at best, a factor of three in regions of high threshold elevation (masking). To predict the data we have developed an adaptive template model where the template matches the stimulus task. This model does a good job of predicting all the characteristics of the data set that includes a broad spectrum of test conditions. We argue that the human visual system, much like and ideal observer model, is able to construct templates for stimuli of this level of complexity under well specified (and rehearsed) tasks. This analysis suggests that standard models with fixed sampling characteristics may be inadequate to predict performance on many tasks.

We have also been applying filter models to more applied stimuli, namely standardized video streams that are used to evaluate digital compression technologies. In a global network, stringent delay requirements for interactive video pose challenges to the conventional, frame-by-frame, synchronous video rendering. The significant delay jitter associated with packet networks is a big problem. To reduce this delay, we have proposed a video coding method, called "delay cognizant video coding (DCVC)" in which each frame is decomposed into separate data flows, which can tolerate different delays. The reconstruction of video at the receiver asynchronously renders the most visually significant information as it arrives. We have been performing basic psychophysical tests on DCVC sequences to estimate the effect of delay on video quality. Careful psychophysical testing has shown that even a single frame delay can be detected, as is expected from the human spatio-temporal sensitivity envelope. However, as we explored the space we observed that for some compressed sequences, DCVC could actually improve subjective image quality. The seven video sequences used in the experiments were standard H.263 test clips. For each sequence there were two delay conditions, no delay and a delay offset of 12 frames (~400 milliseconds) between the low- and high-delay data flows. For each delay condition there was two video compression (MPEG) conditions, high and low. After each presentation the subject was asked to rank order the 4 stimulus conditions for quality using there own subjective criteria.

As expected, higher bit rates delivered better quality. What is more interesting, however, is that within the same compression rate, the sequence with the large delay offset was usually favored over sequences rendered delay jitter free. This very surprising result appears to be related to a blurring of the high temporal frequency components. Traditional computational video quality measures such as mean square error (MSE) and peak signal to noise ratio (PSNR) indicate the introduction of delay would degrade video quality. We also compared the results with predictions from a standard filter based HVS model (MPQM) for these DCVC sequences. As expected, the filter model also predicted that compressed plus DCVC sequences should appear to have slightly poorer quality than for simple compression without DCVC. Here again we see that standard filter models are two simplistic in approach to capture true image quality. The MPQM seems to over estimate high spatio-temporal frequency masking or the significance of these components for perceived image quality as compared to image fidelity.

We believe the reason for the puzzling observation that DCVC can actually improve video quality over compression only conditions, is that regular H.263 compression introduced dynamic noise artifacts are reduced by presenting the same images for several frames in areas where the original video is relatively static. Depending on context, the flickering caused by quantization noise can be very disturbing which DCVC can reduce and thereby improve image quality. To characterize this effect we tried to determine types of scene content that resulted in this improvement of video quality. It's clear that the effect is very context dependent and related to observer expectations of temporal change in different parts of the scene. Evaluation of image quality is much more complicated than we expected, low-level visual masking participates in the perception of image quality but other high level factors including observer expectations are also very important. We designed a battery of simple test stimuli that exhibit the DCVC effects we have noticed which could be used as tests of vision models and compression algorithms that try to take advantage of this effect to improve video quality. The frames in the figure below shows the luminance profiles for two successive frames of a small piece of two blurred circles that are increasing and decreasing in diameter. The change in diameter is about 0.5 min (from row A to row B), which is just barely detectable. The right two images show the luminance profiles of the same two frames after compression. The change in luminance near the base of each ring is very noticeable and disturbing even though the rings themselves appear to be stationary. This is a case where DCVD would greatly enhance the image quality compared H.263. Other stimuli in the battery mimics motion transparency and shadows moving over a texture condition both of with produce similar results. With increased understanding of the impact of dynamic noise on image quality, future video encoders could improve video quality while reducing the bit rate.



Noise masking and adaptable filters in human vision: Noise masking needs to be better understood both for improving our understanding of suprathreshold visual processing and for improving image compression in natural and medical images. Masking by noise is seldom studied in traditional psychophysics yet those studying medical imaging (complex stimuli) commonly use it to study masking. We have been trying to bring together the approaches of multiple disciplines to see how they might bear on the problem of image fidelity. Using noise masking of sinusoidal gratings we have devised several new analytic techniques for overcoming previous limitations in the use of noise for psychophysical testing and have demonstrated the importance of cues that alter stimulus certainty. The results point out the importance of higher order processes in a seemingly low-level visual masking task. There is a strong cognitive component (learning, memory, attention) to the tuning of mechanisms in noise masking.

The usual ideal observer calculation of efficiency ignores the run-to-run fluctuations, which we consider important. For example, in a given run the ideal observer may do poorer than average, because of the particular noise fluctuation and the human observer would likewise do worse than average. The usual calculations would underestimate the efficiency. We calculate ideal observer performance on a trial-by-trial basis to achieve a much more accurate estimate of human observer efficiency. This new method of analysis we hope will become more common in future studies. The following paragraphs goes into these results in more detail since they have yet to be published or discussed in previous reports.

Methods: As in most of our previous studies we utilized the test-pedestal paradigm but in this case add a noise mask. The noise mask was the sum of the first nine harmonics of a 0.5 c/d sinewave fundamental grating. Noise = $a_f \sum_{f=1\text{-}9} \cos(\pi f x) + b_f \sin(\pi f x)$, where a & b are gaussian random numbers. A new noise sample was created for each trial but the a & b coefficients were for stored for later frequency tuning analysis. In all runs different samples of the same noise distribution were used.

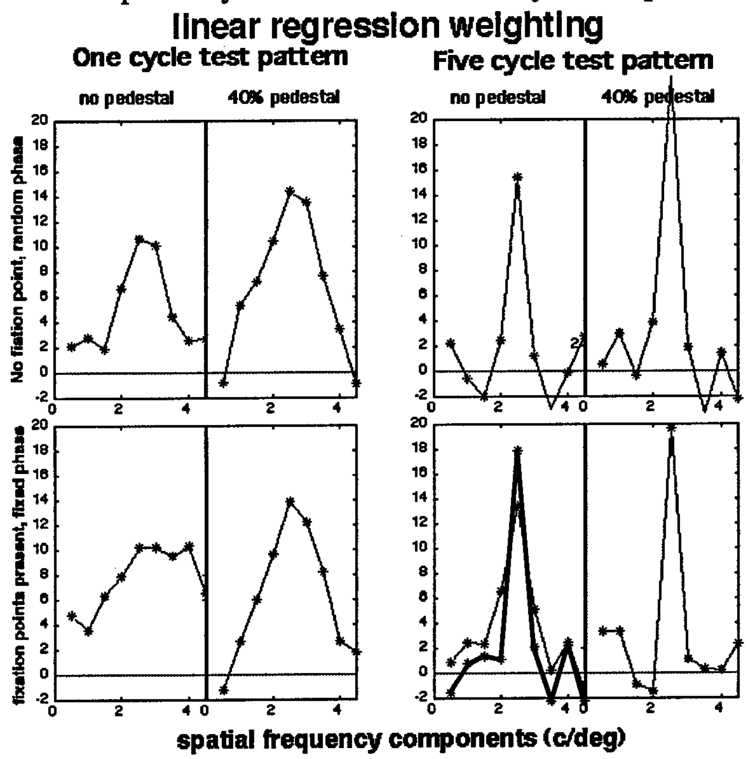
The test+pedestal was a windowed 2.5 c/d sinusoid: test = $(c_p + c_t)\cos(\pi 5x+\theta)((1+\cos(2\pi x))/2)^n$, where c_p and c_t are the pedestal and test contrasts and n = 0 or 4 for the 1 and 5 cycles patterns, respectively. Eight test+pedestal patterns were used with two alternatives for each of three parameters, as follows:

- *Pedestal contrast:* 0% or 40%
- *Number of cycles:* 1 or 5 (enveloped test+pedestal pattern)
- *Phase:* fixed (with fixation marks) or random (no marks)

In each run of 200 trials four test contrasts (0, 1, 2 & 3 times a base contrast) were intermixed. The observer gave a four category rating response corresponding to their estimate of which test contrast was presented. The stimulus duration was 0.3 sec in all but one condition. For the fixed phase, no pedestal, five-cycle test stimulus we did runs at 3.0 sec as well as 0.3 sec. Feedback was provided after each response. The feedback was the calculated ideal observer's response rather than the actual test contrast, on the assumption that it was more reliable since the noise could in some trials actually reduce the test contrast. The ideal observer made judgments on each stimulus trial just as the human observer. How to calculate the idea observer's response is an interesting problem/advance in itself, discussion of which we save for a future publication.

Results: - 1) Determine which frequencies the observer uses: We kept track of the 18 amplitudes of the noise pattern for each trial (9 even and 9 odd components). At the end of each run we modeled both the human and the ideal subject's ratings using the following linear model: rating =

const + $\sum r_f \; c_f$, where c_f are the contrasts of the cosine phase noise components for the phase known case and for the phase unknown case we used the Pythagorean sum of the even and odd components. The r_f coefficients are estimated by linear regression. The two observers had similar results, the figure below shows the average of the two. The upper and lower rows of plots are for the random phase and fixed phase conditions respectively. The data for the one-cycle test patterns (left panels) uniformly show a broad frequency tuning, indicating that the human observer is using the broad range of frequencies that comprise the test pattern. Except for one case, the five cycle test patterns (right panels) show extremely narrow frequency selectivity. That means that in the 0.3 sec exposure the human observers were able to utilize all five cycles of the test pattern. The one exception is the case of no pedestal and fixed phase. In this condition optimal performance requires scrutiny of the locations as well as contrasts of the peaks. For this condition, both observers also did the task with a 3.0 sec duration (shown as bold line). The extra time allowed scrutiny to produce very narrow frequency selectivity.



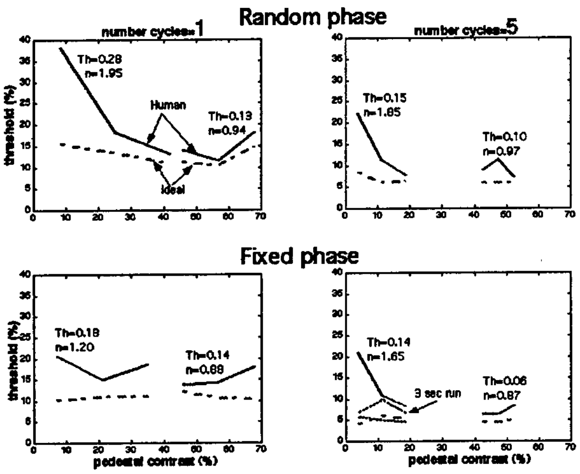
Results: - 2) Discrimination threshold as function of pedestal strength: The two observers' data are combined in the plot below showing an estimate of the TvC function. The abscissa (local pedestal) is the mean contrast between two test levels. The ordinate is the local threshold. The local threshold is obtained by dividing the test increment by the delta d' between adjacent stimulus pairs. This procedure can be justified for pedestal contrasts above threshold where the transducer function becomes linear. However, for simplicity of presentation we use the same formula for the threshold region. That enables us to combine the pedestal =0 and pedestal = 40% into the same panel. For example, suppose the test contrast was 10%. Then we would have points at abscissa values of 5%, 15%, 25%, 45%, 55% and 65%. The human and ideal observer thresholds are identified in the figure. The data for the 3.0 sec condition for the 5 cycle phase known data are also labeled in the figure.

Several items are noteworthy:

- The human shows strong threshold elevation at the leftmost datum for each condition in which the duration was 0.3 sec. This indicates stimulus uncertainty even in the fixed phase condition, where 0.3 sec duration was insufficient to gain certainty. The 3.0 sec duration allowed sufficient scrutiny to bring the human close to the ideal.
- The pedestal greatly reduces uncertainty and the human and ideal curves are very close to each other.

- For both fixed and random phase and both low and high pedestal strengths, the 5 cycle thresholds are substantially lower than the 1 cycle thresholds. There is substantial spatial summation.

Test vs. Pedestal plots in the presence of noise



The overall results have several important implications a few of which are described here.

Spatial summation. Kersten (Vis. Res. 1984) reported that there was negligible spatial summation in the presence of noise. Our results show that thresholds for the five-cycle stimulus were 66% and 63% of the one-cycle thresholds for the no pedestal and 40% pedestal data. These results indicate substantial spatial summation. We believe that the explanation of the difference between our results and Kersten's is that he used temporally white noise, whereas, our noise was static. For the dynamic case it may not help to scrutinize multiple bars to obtain independent samples since one can attend to a single bar to do averaging. In our static case averaging over multiple bars is the only way to do averaging. It is noteworthy that with five cycles our observers had contrast 'efficiencies' of above 50%. The only case of poor efficiency by both observers was for the 5 cycle, fixed phase condition.

Role of scrutiny. By increasing the duration from 0.3 to 3.0 sec Subject 2's human thresholds went from 13.9% to 6.1% while the ideal thresholds went from 3.1% to 4.4%. The efficiency changes from 22.1% to 72%. For Subject 1, human thresholds went from 14.8% to 7.8% while ideal thresholds went from 4.9% to 7.1% corresponding to efficiencies of 33% for 0.3 sec and 91% for 3.0 sec duration. The subjects felt that the three seconds were quite well spent in checking whether there were peaks near the fixation marks.

Run-to-run fluctuations. If one ignores the left-most datum for the human observer where human performance is severely degraded compared to ideal (figure above), then one sees a strong correlation in the fluctuations between the human and ideal observer data. It is useful to keep track of the ideal observer's responses on individual runs when calculating efficiency. Our calculation produces efficiency estimates with smaller standard errors than if the 'absolute' ideal observer's thresholds were used rather than the run-to-run thresholds.

Cognitive components in visual tasks. We developed a quick, reliable method for estimating the frequencies used by subjects in this task. We found that subjects were amazingly efficient at adapting the 'mechanism' bandwidth to the task at hand. This is reminiscent of the earlier results on Vernier acuity, which were best modeled using templates, a kind of adaptive filter. The fixed filter properties of present HVS models will need to include a means of adapting mechanism bandwidths and other characteristics depend on task demands. This is not an easy problem, we first need to determine to what degree and under what conditions the HVS can adapt filter properties.

B) Multitude of masking phenomena in real world situations. (#6, 8, 11, 14, 24)

When we began formulating this research effort about 6-7 years ago the vision science community (us included) focused on two types of masking, contrast gain control and transducer

function with a saturating non-linearity. Our original proposal assumed that once we understood both types, we could build a computational vision model or fidelity metric that could be used to improve image compression. Over the past several years we have learned that the situation is not so clean cut, especially as it applies to real world scenes. While gain control and transducer function masking are important factors in laboratory studies using simple targets, it seems other factors may become just as important, if not more, for complex stimuli in real world applications. Some of these effects have been alluded to in the discussions above involving our experiments with complex stimuli. In our 1997 SPIE presentation we began to realize the magnitude of the problem and identified seven different types of masking. The last two in the list, stimulus uncertainty and intrusive noise, are probably most important in terms of a fidelity metric and are missing from present models (its not clear how to include them as yet).
We are not alone in this general realization, all of the talks in our session of the meeting focused on this very issue. Our studies on noise masking above were designed to start addressing these issues but this is just the beginning of what needs to be researched. The most dramatic demonstration of the failure of present HVS model based fidelity metrics was presented by Dr. Corriveau at this years SPIE meeting (2000). He reported on the VQEG (video quality experts group) international effort to evaluate present video fidelity metrics, eight HVS based metrics and the standard PSNR metric, which uses no information about the human visual system. Performance of the metrics was compared to actual psychophysical measurements on the same natural scene video clips. To everyone's dismay, the PSNR metric performed as well, and in some cases much better than, the HVS based metrics. Clearly, much more research is need before we can successfully extend our models of masking derived from studies using simple stimuli in laboratory setting to real world video content. I think the Modefest group effort is a good start in the direction of using complex stimuli and will become even more relevant as the years progress and they move to increasingly in the direction of more complex targets and tasks.

The 7 general categories of visual masking described in our manuscript are summarized here again for emphasis. Again, the most important of which are categories 6 & 7:

1. Pooled Contrast gain control: This model employs divisive inhibition to set the gain of the optimally responding mechanisms by dividing their activation by the response of other mechanisms that also respond to the test and mask. While this type of model has physiological support it may only account for a small portion of masking in natural scenes.

2. Single component transducer saturation: This is a special case of contrast gain control where the optimal mechanism response is the sum of the mechanism response to the test and to the mask. If the mechanism has a saturating non-linearity then the visibility of the test is a decreasing function of mask strength. Neurons do exhibit a saturation non-linearity. However, as some neurons saturate others with high thresholds are just becoming active.

3. Phase Inhibition or Pythagorean (two component) pooling: Inhibition between phases may be important. Mechanisms with different spatial phases are summed together before a saturating non-linearity stage.

4. Multiplicative noise: The first three categories of masking have the implicit assumption that visual system noise is constant. It's been shown that data fit with a compressive non-linearity transducer function can also be fit with an accelerating transducer function in the presence multiplicative noise.

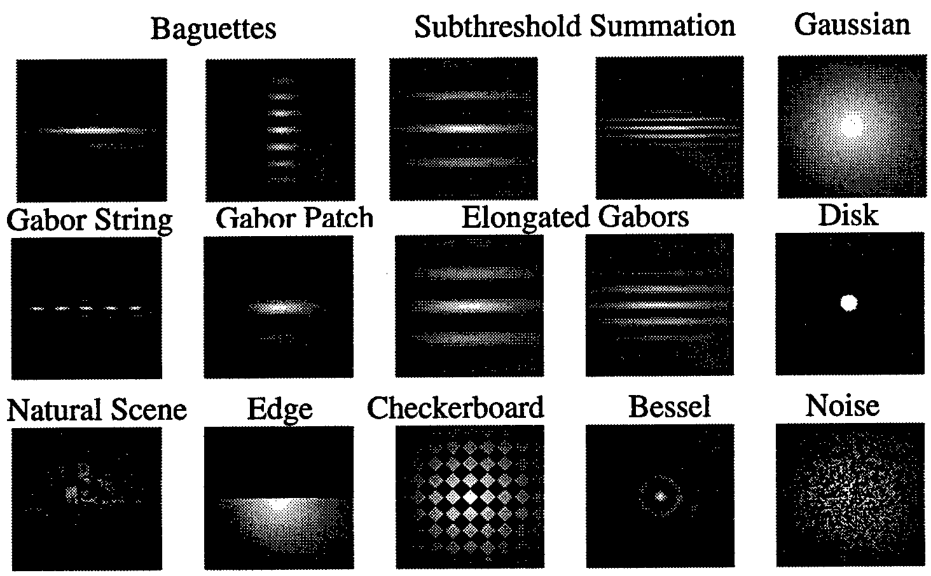
5. Masking by beats (phase locking): The beat pattern associated with amplitude modulated gratings can mask the visibility of a grating at the beat frequency even though the pattern contain no energy at the beat frequency. The masking behavior might reflect a rectifying non-linearity.
6. Stimulus uncertainty: This category along with category seven maybe the most important areas of masking in terms of real world scenes (video segments) yet relatively little research in this area has been performed. Stimulus uncertainty is where the making pattern is confused with the test pattern. Here we are concerned with the cases where the masker noise intrudes because the observer is uncertain about which visual mechanism contains the signal. In this category we have three sub-categories: *Imperfect memory*: In many cases the mask appearance must be remembered as in a contrast discrimination task. If response to the mask alone is not remembered correctly then it could be confused with stimulation by the test stimulus alone. This is an important condition in image fidelity of video sequences which would typically include a large memory component when mask (original image) and mask plus test (codec image) image sequences are compared sequentially. *Mistaken identity*: Here the mask and test are activating different visual mechanisms but the observer confuses the two classes of responding mechanisms. For example, discrimination of two overlapping sinewaves (test and mask) of similar frequency with spatial phase and amplitudes randomized over trials. Phase uncertainty would greatly elevate the discrimination threshold. If phase of the test is unknown this elevates the transducer function slope. The final decision rule is critical, how does the system combine information across activated mechanisms to discrimination test from mask. *Pooling across mechanisms or complex cell pooling*: This category is similar to the previous except here we think of the pooling across mechanisms to be hard wired, much like the integration seen at the complex cell stage of the visual cortex. Here the decision strategy is based on a weighted pool of mechanisms. A simpler but less efficient ideal observer rule that might be used here for the phase uncertainty issues described above.
7. Intrusive noise: In this category the intrusive noise directly contributes to the mechanism detecting the test. In the previous category the intrusive noise tended not to overlap with the test (it did not directly activate the mechanisms detecting the test), the intrusion resulted from cognitive uncertainty between mechanisms responding to the test and mechanisms responding to the mask. Both types of intrusive noise will likely have a large effect on detecting a test pattern in the presence of a pedestal or mask pattern. This source of masking will have its largest effect in a complex visual scene where stimulus uncertainty will be maximized. In simple visual tasks, as commonly used in vision science experiments to reveal underlying mechanism function, the stimuli tend to minimize intrusive noise masking. At present the models designed for use as a fidelity metric do not incorporate intrusive noise masking and therefore are likely to greatly underestimate masking, especially in images of a complex scene or video sequences.

C) Modelfest - an innovative approach to vision modeling (#20, 22, 23, 25, 26, 27, & 28)

Over the past 35 years, the vision science community has made significant progress in understanding the early stages of visual processing. Visual psychophysics and physiological studies have revealed a multi-stage parallel processing structure of the early human visual system (HVS). Although most HVS models exhibit similarities, such as banks of Gabor filters, they have distinct differences in how they combine filter responses and account for visual masking. Are the model differences significant? Under what conditions does one model perform better than another? These questions are very hard to answer because models are rarely compared using the

same psychophysical data set. As a result, the efficacy of different models is unclear. At a modeling workshop in the 1997 annual OSA meeting, we began setting the framework for the Modelfest group to address the issues and enhance cross-fertilization among vision modelers through several means. The general feeling at this and subsequent meetings was we need to 1) develop a large public database of psychophysical thresholds with stimuli designed to challenge and facilitate the development of HVS models. 2) devise a scheme for evaluating models using the public database of stimuli and psychophysical thresholds. As models become complex, comparisons of their efficacy can be very difficult. 3) through the public meetings of the Modelfest group, can foster the sharing of ideas between members and encourage other vision researchers to contribute to HVS modeling. Previously, there has been minimal cross-fertilization among vision modelers, and finally, 4) provide a "standard observer" data set for spatio-temporal vision, much like color vision has had for many decades.

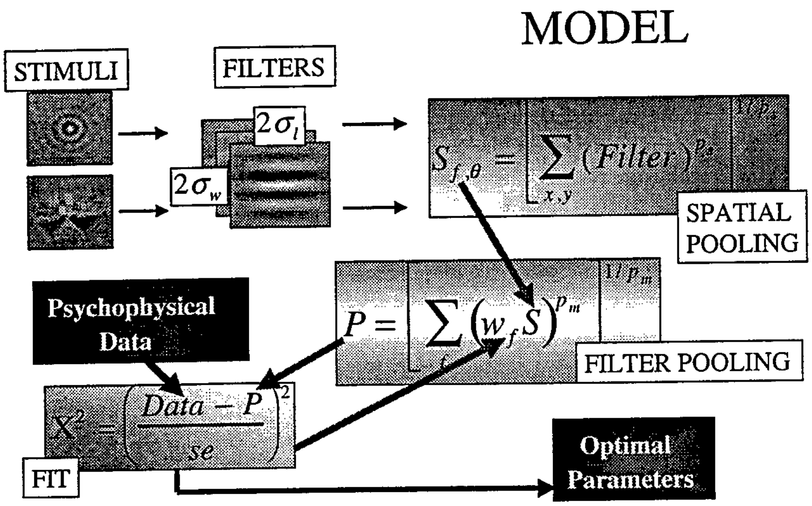
In June 1998, I organized, and continue to administer, the Modelfest data collection group. The 12-member data collection group devises stimuli that are deemed critical for developing and challenging vision models. Our goal is to provide an extensive public stimulus database to be used for testing different aspects of HVS models. The database includes psychophysical threshold data, from laboratories across the country, on each of the stimuli in the database. The database will grow each year, and presently contains detection thresholds for a set of 44 stimuli. The figure below demonstrates the variety of targets in the database. All the stimuli and preliminary data from the first year's data collection effort are posted: www.neurometrics.com/ .



Future years will include stimuli designed to challenge models in areas such as contrast masking, and color vision. Once a large, readily accessible database of stimuli with psychophysical thresholds exists, the developers of general purpose HVS models will be compelled to provide performance data using the database images before their model will be taken seriously. As more models are applied to this common data set it will become much easier to determine which model innovations actually improve model performance. Modelfest is a dramatic change from how HVS modeling has progressed in the past. This exciting new approach offers the field a simple way of comparing models and learning from each other's innovations and mistakes.

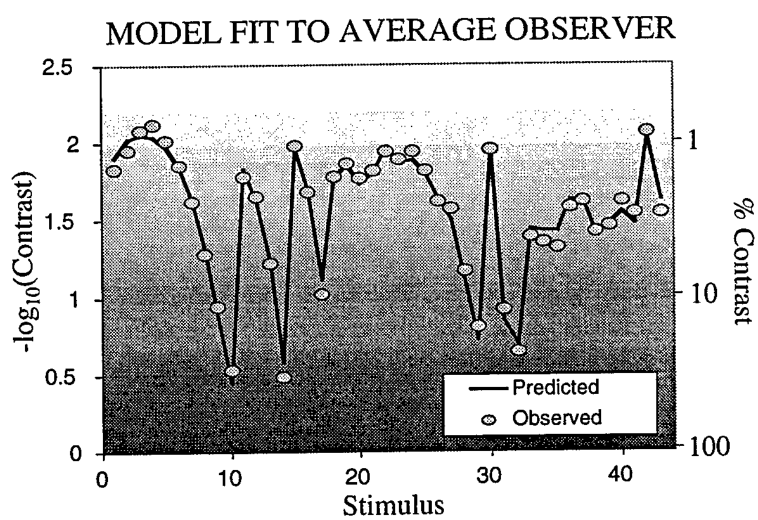
Several groups have already started modeling the first Modelfest dataset, ourselves included. We have designed a spatial vision model based on common assumptions about early visual mechanisms. The goal was to see how well it predicts the database thresholds. Moreover, we wanted to determine how well the test battery of stimuli constrained model parameters. Our

model incorporated 7 free parameters: spatial and filter pooling exponent parameters, three contrast sensitivity function parameters and two filter bandwidth parameters. The figure below outlines the model showing the steps performed to determine optimal parameter values.



The model predictions were very accurate as shown in the figure below. The average group threshold for each of the 43 stimuli is plotted with open symbols. The solid line indicates the predicted threshold using the optimal model parameters. The prediction error for each stimulus was within 1 db. Filter bandwidth was tightly constrained at 1.5 octaves but filter length tuning was poorly constrained.

This version of the 'standard model' can accurately fit the Modelfest dataset. However, the Modelfest dataset does not appear to adequately cover the stimulus space to constrain the filter length in this version of the 'standard model'. We are now extending the dataset to find stimuli that better fix the filter bandwidth parameters. While the success of the model is impressive given the diversity of stimuli, this was only based on detection thresholds and avoids the topic of masking. Future Modelfest data collection efforts will focus on masking which will certainly pose a serious modeling challenge.



One of the goals of Modelfest was to facilitate interactions between members. The fruits of this approach have already been demonstrated to us, Modelfest members have pointed out to us limitations of the function we used to characterize the CSF. I'm sure using a recommended CSF model we could improve upon the fit shown in the figure above. Modelfest offers great promise as a vehicle to enhance future modeling efforts by providing standardized datasets and facilitating interactions between laboratories across the country.

AFOSR supported manuscripts published or submitted and formal presentations:

Manuscripts that are underlined are included along with this final report as appendices. Other publications were submitted along with earlier annual reports.

1. S. A. Klein, T. Carney and Q. Hu, (1995) Improved lookup table to correct CRT adjacent pixel nonlinearity. *Proceedings of the SPIE* "Human Vision, Visual Processing, and Digital Display VI". Ed: Rogowitz, B. E. & Allenbach, J. P. **2411**, 170-179.
2. T. Carney, S. A. Klein, and Q. Hu, (1995) Asymmetries of visual masking for the detection of light and dark bars. *OSA Digest* annual meeting, Portland, Oregon p64.
3. S. A. Klein, B. L. Beard, and T. Carney (1995) Motion mechanisms and contrast mechanisms have similar saturation behavior. *OSA Digest*, Portland, Oregon p64.
4. T. Carney, S. A. Klein, and Q. Hu, (1996) Visual masking near spatiotemporal edges. *Proceedings of the SPIE* "Human Vision, Visual Processing, and Digital Display VI". Ed: Rogowitz, B. E. & Allenbach, J. **2657**, 393-402.
5. T. Carney and S. A. Klein (1996) Moving resolution acuity from the test-pedestal perspective. *Investigative Ophthalmology & Visual Science* **37**, S747.
6. S. A. Klein and T. Carney (1996) Transducer function for energy (viewprint) models. *Society for Information Display 96 Digest XXVII.* **27**, 425-428.
7. S. A. Klein, Q. Hu and T. Carney (1996) The adjacent pixel nonlinearity: Problems and solutions. *Vision Research* **36**, 3167-3182.
8. S. A. Klein, and T. Carney (1996) A steep ideal observer transducer function for detecting phase unknown sinusoids. *OSA annual meeting Digest*, Rochester, NY.
9. T. Carney and S. A. Klein (1997) Resolution acuity is better than Vernier acuity *Vision Research* **37**, 525-540.
10. T. Carney and S. A. Klein (1997) Gray scale adjacent pixel luminance non-linearity compensation with 3 color guns. *Proceedings of the SPIE* "Color Imaging: Device-Independent Color; Color Hard Copy and Graphic Arts II". Ed: Giordano B. Beretta & Reiner Eschbach, **3018**, 188-197.
11. S. A. Klein, T. Carney, L. Barghout & C. Tyler (1997) Seven models of visual masking. *Proceedings of the SPIE* "Human Vision, Visual Processing, and Digital Display". Ed: Rogowitz, B. E. & Allenbach, J. P., **3016**, 13-24.
12. T. Carney (1997) Evidence for an early motion system which integrates information from the two eyes. *Vision Research* **37**, 2361-2368.
13. B. L. Beard, S. A. Klein & T. Carney (1997) Motion thresholds can be predicted from contrast discrimination. *Journal of the Optical Society of Amer. (A)* **14**, 2449-2470.
14. S. A. Klein, T. Carney, L. Barghout-Stein & C. Tyler (1997) Questioning gain control models in psychophysics. *OSA Annual Meeting: Suppl. Optics & Photonics News* **8,** 68.
15. T. Carney (1997) Tools for modeling and testing visual function. *OSA Annual Meeting: Suppl. Optics & Photonics News* **8,** 85.
16. D. M. Levi, S. A. Klein and T. Carney (1998) Unmasking multiple mechanisms for Verner acuity. *Investigative Ophthalmology & Visual Science* **39**, S411.
17. M. Levi, S. A. Klein, & T. Carney (1998) Orientation, spatial frequency and length tuning in vernier acuity. *Annual ECVP meeting.*

18. Y, Chang, T. Carney, S. A. Klein, D. G. Messerschmitt and A. Zakhor (1998) Effects of temporal jitter on video quality: assessment using psychophysical methods. *Proceedings of the SPIE* "Human Vision and Electronic Imaging III". Ed: Rogowitz, B. E. & Allenbach, J. P. **3299**, 173-179.
19. T. Carney and S. A. Klein (1999) Optimal spatial localization is limited by contrast sensitivity. *Vision Research* **39**, 503-511.
20. Y, Chang, D. G. Messerschmitt, and T. Carney (1999) Expanding Network Video Capacity with Delay-Cognizant Video Coding. *Proceedings of the SPIE "Visual Communications and Image Processing 99"* Ed. Aizawa, K., Stevenson, R.L. & Zhang, Y. **3653**, 783-794.
21. T, Carney, S. A. Klein, C. W. Tyler and A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C. Chen, W. Makous & M. P. Eckstein (1999) The development of an image/threshold database for designing and testing human vision models. *Proceedings of the SPIE* "Human Vision and Electronic Imaging IV". Ed: Rogowitz, B. E. & Pappas, T., **3644**, 542-551.
22. T. Carney, Y. Chang, S. A. Klein and D. G. Messerschmitt (1999) Effects of dynamic quantization noise on video quality. *Proceedings of the SPIE* "Human Vision and Electronic Imaging IV". Ed: Rogowitz, B. E. & Pappas, T.N., **3644**, 141-151.
23. T. Carney (1999) Special Interest Group, Modelfest: Vision modeling - Progress and future plans. *Investigative Ophthalmology & Visual Science (supp)* **40**, Insert-9.
24. S. A. Klein and T. Carney (1999) Unmasking noise masking: Spatial summation in detection and discrimination. *Investigative Ophthalmology & Visual Science,* **40**, S43.
25. S. A. Klein, (1999) "Modelfest '99 Workshop: Comparing detection models," *Optical Society of America Annual Meeting, Digest of Technical Papers* pp. SuE., 1999.
26. L. Walker, S. A. Klein, and T. Carney, (1999) Modeling the Modelfest data: decoupling probability summation, *Optical Society of America Annual Meeting, Digest of Technical Papers*, pp. SuC5.
27. T. Carney, L. Walker, S. A. Klein, (2000) "Multi-scale spatial detection model prediction of the Modelfest dataset," *Investigative Ophthalmology and Visual Science* **41**, (inpress).
28. T. Carney, C. W. Tyler, A. B. Watson, W. Makous, B. Beutter, C-C. Chen, A. M. Norcia, & S. A. Klein (2000) Modelfest: year one results and plans for future years *Proceedings of the SPIE* "Human Vision and Electronic Imaging V". Ed: Rogowitz, B. E. & Pappas, T. **3959**, in press.
29. D. M. Levi, S. A. Klein and T. Carney (2000) Unmasking the mechanisms for Vernier acuity: evidence for a template model for vernier acuity. *Vision Research.* (in press)
30. D. A. Silverstein, T. Carney and S. A. Klein (submitted, 2000) Modeling contrast thresholds. *Book chapter. Pergamen Press*
31. Y, Chang, D. G. Messerschmitt, T. Carney, and S. A. Klein (submitted, 2000) Delay Cognizant Video Coding: Architecture, Applications and Quality Evaluations. *IEEE Transactions*

Professional personnel:

Thom Carney, PhD. Principal Investigator

Stanley A. Klein, PhD.

Interactions: Presentations at meetings (final year):

Optical Society annual meeting 1999

S. A. Klein, (1999) "Modelfest '99 Workshop: Comparing detection models," *Optical Society of America Annual Meeting, Digest of Technical Papers* pp. SuE., 1999.

L. Walker, S. A. Klein, and T. Carney, (1999) Modeling the Modelfest data: decoupling probability summation, *Optical Society of America Annual Meeting, Digest of Technical Papers*, pp. SuC5.

ARVO annual meeting 1999-2000

T. Carney (1999) Special Interest Group, Modelfest: Vision modeling - Progress and future plans. *Investigative Ophthalmology & Visual Science* **40**, Insert-9.

S. A. Klein and T. Carney (1999) Unmasking noise masking: Spatial summation in detection and discrimination. *Investigative Ophthalmology & Visual Science,* **40**, S43.

T. Carney, L. Walker, S. A. Klein, (2000) "Multi-scale spatial detection model prediction of the Modelfest dataset," *Investigative Ophthalmology and Visual Science* **41**, (in press).

SPIE annual meeting 1999-2000

T, Carney, S. A. Klein, C. W. Tyler and A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C. Chen, W. Makous & M. P. Eckstein (1999) The development of an image/threshold database for designing and testing human vision models. *Proceedings of the SPIE* "Human Vision and Electronic Imaging IV". Ed: Rogowitz, B. E. & Pappas, T., **3644**, 542-551.

Y, Chang, D. G. Messerschmitt, and T. Carney (1999) Expanding Network Video Capacity with Delay-Cognizant Video Coding. *Proceedings of the SPIE "Visual Communications and Image Processing 99"* Ed. Aizawa, K., Stevenson, R.L. & Zhang, Y. **3653**, 783-794.

T. Carney, Y. Chang, S. A. Klein and D. G. Messerschmitt (1999) Effects of dynamic quantization noise on video quality. *Proceedings of the SPIE* "Human Vision and Electronic Imaging IV". Ed: Rogowitz, B. E. & Pappas, T.N., **3644**, 141-151.

T. Carney, C. W. Tyler, A. B. Watson, W. Makous, B. Beutter, C-C. Chen, A. M. Norcia, & S. A. Klein (2000) Modelfest: year one results and plans for future years *Proceedings of the SPIE* "Human Vision and Electronic Imaging V". Ed: Rogowitz, B. E. & Pappas, T. **3959**, in press.